Phonological factors in Tagalog adjective-noun word order variation*

Stephanie S Shih
Stanford University/UC Berkeley
stephsus@stanford.edu

Kie Zuraw
UCLA
kie@ucla.edu

1 INTRODUCTION TAGALOG ADJECTIVE/NOUN PAIRS

- Both orders are possible and claimed to be semantically interchangeable = "without any apparent difference in meaning" (Schachter and Otanes 1972: 122):

(1)	magandá-ng	babáe	~	babáe-ng	magandá	(Kroeger 1998)
	beautiful-LINK	woman		woman-LINK	beautiful	'beautiful woman'

- Elsewhere, it's been shown that word order variation can be conditioned by numerous factors
  - phonological, syntactic, semantic, usage-based, psycholinguistic, sociolinguistic (e.g., McDonald et al. 1993; Wasow 2002; Wright, Hay & Bent 2005; Benor & Levy 2006; Bresnan et al. 2007; Anttila et al. 2010; and references therein).

❖ **What factors (if any) condition Tagalog adjective/noun word order variation?**

- An additional wrinkle:
- The "linker" morpheme, attached to first word, has phonologically conditioned allomorphs:

(2)	a. If word ends in V, add *ng* ([-ŋ]):	babáe → babáe-ng	'woman'
	b. If word ends in /n/ or /ʔ/, change it to *ng* ([ŋ]):	karaníwan → karaníwa-ng	'ordinary'
	c. If word ends in any other C, add [na]:	itím → itím na	'black'

❖ **Do factors that condition word order variation depend on the phonological (surface) form of the linker?**

1.1 THIS TALK

- Large-scale corpus study investigating factors affecting adjective/noun order.
- As far as we know, this is the first systematic study of Tagalog adjective/noun order variation.
- In particular, we focus on phonological factors that may affect order, including segmental and prosodic well-formedness conditions.

* Acknowledgments to RAs Seth Ronquillo, Arienne Filio, and Kimberly Nachor for data checking; Vera Gribanova, Laura McPherson, and participants in the UCLA Phonology Seminar for comments. Ivan Tam developed the software that created the Tagalog corpus.

Handout available http://stanford.edu/~stephsus/ShihZuraw-LSA2014.pdf

Results

- In addition to semantic and usage-based factors, phonological conditions influence adjective/noun word order in Tagalog.
- Word order is conditioned in part by the phonological surface form of the linker particle.

  ↳ Word order can optimize phonological well-formedness (see also Schlüter 2005; Shih, in prep).
  ↳ Phonological (surface) information is available at the point of word order choices.

.

- Theoretical implications of our results:
  - for understanding the role of phonology in determining word order, and
  - for considering the design of the interface between phonology, morphology and syntax in both formal grammatical models and psycholinguistic models of language production.

2 DATA

2.1 TAGALOG WEB TEXT CORPUS FROM (ZURAW 2006)

- Web text from 2004, variety of genres
- 47,144,971 word tokens, 105,464 word types

2.2 ADJECTIVE/NOUN PAIR EXTRACTION

- Nouns and adjectives automatically and manually extracted from part-of-speech tags on the SEAsite online Tagalog-English dictionary (SEAsite 2001).
- Searched corpus for all possible noun-linker-adjective and adjective-linker-noun sequences.
- Automatic exclusion of pairs in the following circumstances:
  - instances wherein the second word potentially had a *ng* linker
  - instances with punctuation occurring between the two words
  - instances where a noun or adjective must bear a certain affix according to the Seasite Dictionary but does not
  - instances with the following tokens: *alám*, *am*, *dápat*, *habang*, *hanggáng*, *lamáng*, *saíd*, *sayá (ng)*, *tápos*, *tódo(ng)*, *úpang*[1]
- Adjective/noun pairs hand-checked by three Tagalog-English bilinguals:
  - 1,205 noun/adjective pairs (types) were selected, to include the nouns and adjectives occurring most frequently in the set, as well as the most-frequent pairs.
  - 11 words identified as problematic: adverbs rather than adjectives; ambiguities caused by linker (e.g., *noóng* could be *noó-ng* 'forehead' + linker, or *noóng* 'when-*past*').
  - Adjective/noun pairs containing any problematic words were excluded.
- Words' stress patterns manually obtained from a paper dictionary (English 1986).

[1] *Am* (e.g., versus PM) and *said* were excluded as English words. The rest are more commonly used as verbs (*alam*, *dapat*), prepositions (*habang*, *hanggang*, *tapos*, *upang*), quantifiers (*todo*), or enclitics (*lamang*, *saya*).

- 149,689 adjective/noun pair tokens, 14,591 types
- 1,708 noun types, 587 adjective types

- Some caveats = potential sources of noise:
  - *X*-linker-*Y* might not form a constituent: *Y* may be part of a complex modifier.

(3) a. ***táo-ng armádo*** 'person-LINK armed'
b. ***táo-ng armádo*** *sa kanilang ika-apat na henerasyon mobile na aparato*
'person- LINK armed with their fourth LINK generation mobile LINK device'

  - We don't know which pairs are predicates and which are not (claimed to be important by Schachter and Otanes 1972).
  - We don't know if either of the words is focused or represents new information, etc. (though we hope to add givenness information in future versions of the corpus)s

3 METHODOLOGY

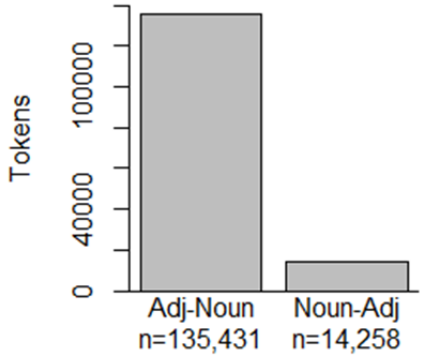
- Mixed-effects regression model using `glmer()` from `lme4` R package (Bates et al. 2013; R Core development team):
  - Multivariate analysis controls for numerous predictors at once.
  - Data: unique adjective/noun combinations (n = 14,591)
  - Dependent variable: rate of noun-adjective order
    = (Instances in noun-adjective order) / (Total instances in both orders)
  - Predictors were centered; numerical predictors also standardized following (Gelman & Hill 2007; Gelman 2008).
  - Random intercepts: noun, adjective
- (We are also working on a binomial model of adjective/noun instances, but results are currently not yet ready to be reported due to model convergence details.)

- Independent variables, culled from previous studies of word order variation, descriptions of Tagalog adjective/noun word order, Tagalog phonology, cross-linguistically common phonological behaviors (see §5).
  - prosodic phonological predictors
  - segmental phonological predictors
  - non-phonological predictors

4 GENERAL RESULTS

- Adjective-noun order is overwhelmingly preferred = "default" order?

(4)



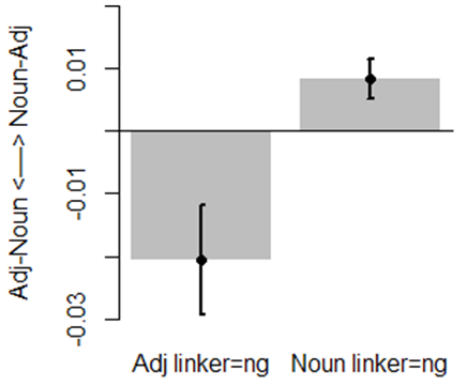
- adjective-noun order = 135,431 tokens
- versus noun-adjective order = 14,258 tokens

- Linker preference: *-ng* linker preferred to *na*.
  - (Schachter and Otanes 1972): If one order requires *-ng* linker and the other requires *na* linker, order that results in *-ng* linker is preferred.
  - Especially true for *-ng* linker adjectives.

(5)

a.	áso-**ng**	ulól	*is more frequent than*	ulól	**na**	áso	
	dog-LINK	mad		mad	LINK	dog	'mad dog'
b.	bágo-**ng**	títser	>	títser	**na**	bágo	
	new-LINK	teacher		teacher	LINK	new	'new teacher'

(6)



- Regression coefficient estimates:

  Adjective linker = *ng*
  (β = -0.02065, SE = 0.004365, t = -4.73)
  Noun linker = *ng*
  (β = 0.00839, SE = 0.001621, t = 5.18)

(7)

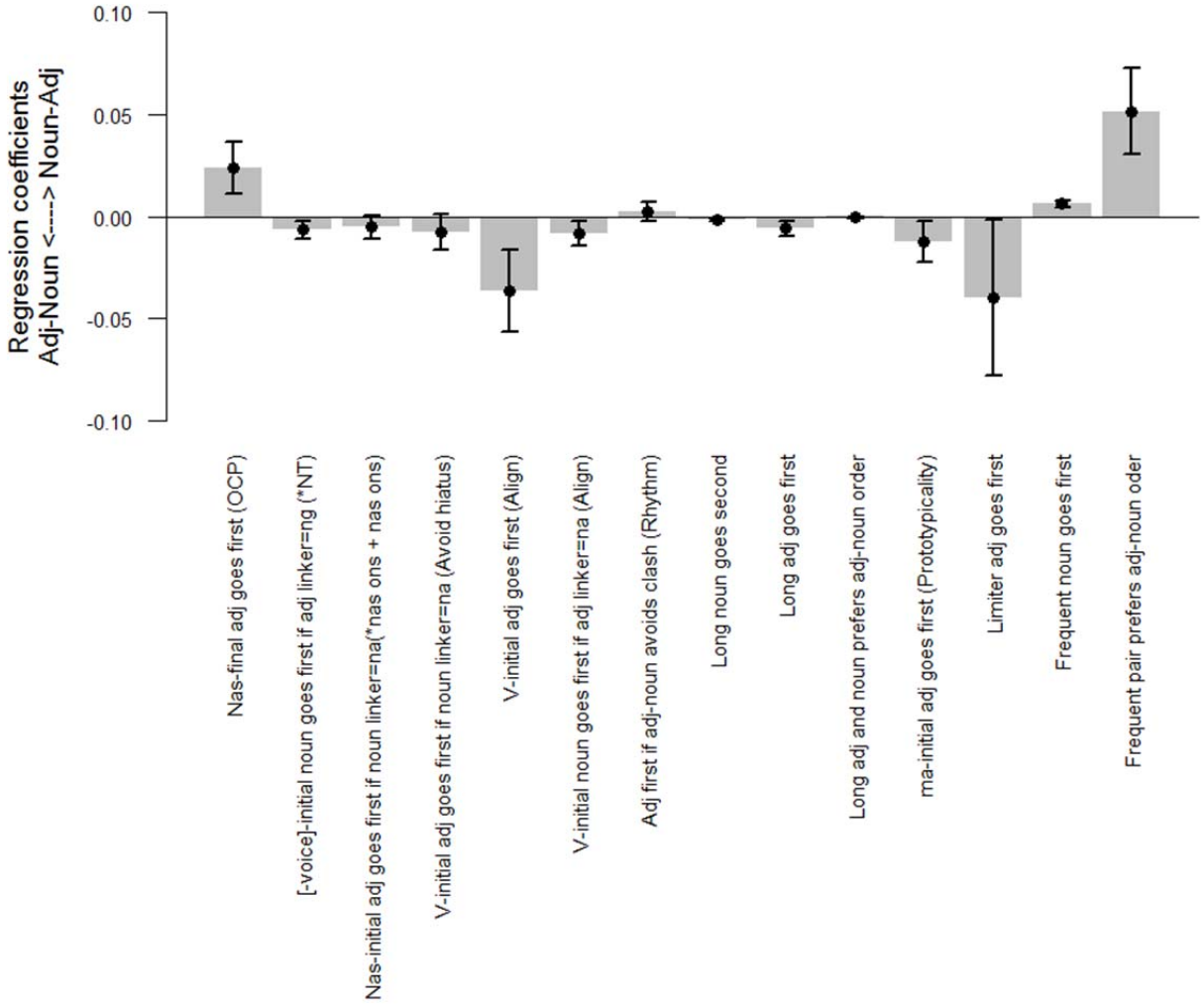
	Phonological predictors tested	*Effect?*
a.	Phonotactic, Obligatory Contour Principle	
	*[nasal] + [nasal]	✓
	*[velar] + [velar]	×
b.	Contextual markedness (e.g., * NÇ)	✓
c.	Long-distance phonotactics (e.g., *nasal onset + nasal onset)	(✓)
d.	Phonological faithfulness (e.g., avoid replacing [ʔ] with nasal)	×
e.	Hiatus avoidance (e.g., *V+V) and syllable structure optimization	(✓)

f.	Phonological and morphological alignment	✓
g.	Stress lapse avoidance	(✓)
h.	Stress clash avoidance	(✓)
i.	Length in segments of adjective and noun	✓
j.	Linker surface form (*-ng* versus *na*)	✓

(8)

	Non-phonological predictors tested	*Effect?*
a.	Adjective prototypicality (*ma-* prefix)	✓
b.	Quantifier/limiter adjectives	✓
c.	Total length in segments	✓
d.	Log frequency: word, pair (calculated from Tagalog Web Text Corpus (Zuraw 2006))	✓

(9) Regression coefficients of reliable and trending predictors (see details in §5)



5 RESULTS: PREDICTORS OF ADJ/NOUN WORD ORDER

5.1 SEGMENTAL PHONOLOGY RESULTS

5.1.1 Obligatory Contour Principle effects

- Avoidance of adjacent similar segments affects genitive construction choice in English: adjacent sibilants are avoided (Menn & MacWhinney 1984; Zwicky 1987; a.o.).

(10) Genitive alternation (Hinrichs & Szmrecsányi 2007; a.o.)

the descendants of the veterans > the veteran**s** + **'s** descendants

- Root-internally in Tagalog, nasal-nasal sequences are underrepresented:
  - In a database of 4,617 native, disyllabic roots (from English 1986), of which 1,579 have a medial cluster, we expect 51 nasal-nasal medial clusters, but find only 8.

➢ Tagalog word order: *[nasal] + [nasal] sequences involving [m, ŋ] + *na* linker are dispreferred.
↳ Avoid [m, ŋ]-final adjective in first position, which takes linker *na*.
(β = 0.023796, SE = 0.006402, t = 3.72)

(11)

tsinélas	**n**a	itím**m**	>	itím**m**	**n**a	tsinélas	
slippers	LINK	black		black	LINK	slippers	'black slippers'

*5.1.2 Contextual markedness: *NÇ*

- Cross-linguistically, NÇ clusters can trigger a number of phonological repairs (e.g., deletion, assimilation, fusion; Pater 1996, 2001).
- The *NÇ constraint (Pater 1996, 2001; Hayes and Stivers 1996) is active at Tagalog prefix-stem boundaries, as in many related languages (see Zuraw 2010 for more):

(12) A stem-initial voiceless obstruent usually fuses with a preceding nasal, but a voiced obstruent usually does not.

a. /ma-pa**ŋ-k**amkám/ → [ma-pa-**ŋ**amkám] 'rapacious'
b. / pa**ŋ-d**iníg/ → [pan-diníg] 'sense of hearing'

➢ Tagalog word order: Avoid linker *-ng* + voiceless-initial noun.
(β = -0.006461, SE = 0.002127, t = -3.04)

(13)

péra-ng	nakalaán	>	nakalaá-**ng**	**p**éra		
money-LINK	dedicated		dedicated-LINK	money	'dedicated money'	

5.1.3 Long-distance phonotactics

- Non-local onset similarity avoidance is perhaps a bit more unusual, but repeated similar or identical sequences are often avoided (Stemberger 1981; Menn & MacWhinney 1984; Yip 1998; Frisch et al. 2004; Löfstedt 2010).

(14) Center-embedding of possessive NPs in Ancient Greek blocked when a sequence of adjacent identical articles would be formed (Golston 1995: 353).

a. [[**h-ee** tólm-a] [**t-óon** leg-ónt-oon]]
the-N:F courage-N:F the-G:M:P speak-ing-G:M:P
'the courage of those speaking'

b. ✓[[**h-ee** [**t-óon** leg-ónt-oon] tólm-a]
the-N:F the-G:M:P courage-N:F speak-ing-G:M:P
'the courage of those speaking'

c. [[[**t-óon** oikeí-oon] tin-às] [**t-óon** ekeín-oon]]
the-G:M:P slave-G:F:P some-A:F:P the-G:M:P those-G:M:P
'some of the slaves of those [people]'

d. *[[[**t-óon** [**t-óon** ekeín-oon] oikeí-oon] tin-às]
the-G:M:P the-G:M:P those-G:M:P slave-G:F:P some-A:F:P
'some of the slaves of those [people]'

➢ Tagalog word order: trend of onset similarity avoidance between *na* linker and following N-initial adjective.
↳ Avoid linker *na* + N-initial adjective.
(β = -0.005072, *SE* = 0.002816, t = -1.8)

(15) **m**aóng **n**a kupás > kupás **n**a **m**aóng
denim LINK faded faded LINK denim 'faded denim'

5.1.4 Syllable structure effects

- Hiatus avoidance and ONSET (Prince & Smolensky 1993:17) affects binomial ordering of name pairs in English (Wright et al. 2005; Benor and Levy 2006; a.o.)

(16) John and Yoko > Yok**o** **a**nd John

➢ Tagalog word order: trend of hiatus (and resulting [ʔ] epenthesis) avoidance in noun-adjective order:
↳ Avoid *na* linker + V-initial adjective.
(β = -0.007616, *SE* = 0.004399, t = -1.73)

(17) **e**spesyál na bágay > bágay n**a** ʔ**e**spesyál
special LINK thing thing LINK ordinary 'special thing'

5.1.5 Phonological and morphological alignment

- Morpheme boundaries prefer to coincide with syllable boundaries.
  - e.g., (McCarthy & Prince 1993) ALIGN(Stem, Right; Syllable, Right) for Axininca Campa, Lardil, Hebrew, Bedouin Arabic, and Kamaiurá
  - Alignment between syllable boundaries and higher-level boundaries can affect ordering of syllables in these cases as well: e.g., choice between [na.ta] and *[ta.na] in Axininca Camp depends on ALIGN-L.

➢ Tagalog word order: V-initial words prefer to be initial, avoiding resyllabification or glottal stop insertion.

↳ Avoid V-initial adjectives in second position.
(β = -0.036161, *SE* = 0.010007, t = -3.61)

(18)

ordináryo-ng	táo	>	táo-ng	**o**rdináryo	
ordinary-LINK	person		person-LINK	ordinary	‘ordinary person’

↳ Avoid *ng* linker + V-initial noun.
(β = -0.008055; *SE* = 0.003014, t = -2.67)

(19)

itlóg	na	pulá	>	pulá-**ng**	**i**tlóg	
egg	LINK	red		red-LINK	egg	‘red [brined] egg’

5.2 PROSODIC PHONOLOGY RESULTS

5.2.1 Rhythm

- Stress clashes are avoided via alternative word order in English (see also Temperley 2009):

(20) Clash avoidance leads to avoidance of *a*-adjectives in prenominal positions in diachronic change in English (Schlüter 2005).

a. **σ́** σ σ σ σ **σ́** σ **σ́** **σ́** σ
the **pér**son who was a**wáre** > ?the a**wáre pér**son

b. σ **σ́** **σ́** σ
> ??the a**sléep pér**son

- Stress lapses are avoided via alternative word order in English:

(21) Genitive alternation (e.g., Shih et al. to appear; Shih in prep; see also binomial pair ordering: McDonald et al. 1993; Wright et al. 2005; Benor & Levy 2006)

σ́ **σ** σ́ σ σ		σ́ **σ σ σ σ** σ́ σ	
the sýstem's bénefits	>	the bénefits of the sýstem	

- Background: Tagalog primary stress is contrastive: can be final, penultimate, or in some loans, antepenultimate.
- Stressed non-final vowels are notably longer than unstressed.
- Secondary stress is not fully understood for Tagalog (French 1988; French 1991; a.o.).
  - Usually not included in dictionary; we did not attempt to annotate for it

➢ Tagalog word order: trends in data towards clash and lapse avoidance in adjective-noun word order. The trends are evidenced in a subset of the data controlling for word length (since rhythmic measures are correlated with word length).

➢ Tagalog word order: primary stress clashes caused by adjective-noun order are avoided.
(in data subset with bi- and tri-syllabic words only: $\beta = 0.002865$, $SE = 0.002732$, $t = 1.049$)

(22)

σ́ σ	σ σ́	>	σ **σ́**	**σ́** σ			
báya-ng	sawí?	>	sa**wí-ng**	**bá**yan	'unfortunate		
country-LINK	unfortunate		unfortunate-LINK	country	country/people'		

➢ Tagalog word order: long primary stress lapses (3 or more unstressed syllables) are avoided in adjective-noun order.
(in data subset with bi- and tri-syllabic words only: $\beta = -0.004239$, $SE = 0.002431$, $t = -1.744$)

(23)

σ σ σ́ σ	σ́ σ	>	σ́ **σ σ**	**σ σ** σ́				
espesyál	na	áraw	>	á**raw**	**na**	**espe**syál		
special	LINK	day		day	LINK	special		'special day'

5.2.2 Weight/length

- Heavier/longer constituents tend to come at the peripheries (Behagel 1909; Quirk et al. 1985; Hawkins 1994; Wasow 2002; a.o.).
  - In English, heavier/longer constituents come last. In V-final languages like Japanese, heavier-longer constituents tend to come first (e.g., Hawkins 1994: 426).
  - Numerous proposals for the underlying cause of heavy-to-edge phenomena include processing-based (e.g., Gibson 2000; Temperley 2006), syntactic (e.g., Wasow 2002), and prosodic explanations (Zec and Inkelas 1990; Zubizarreta 1998; Anttila et al. 2010; a.o.) (see Szmrecsányi 2004; Grafmiller & Shih 2011 for empirical comparisons of these proposals).

(24) "Heavy-last" in genitive syntactic variation in English (e.g., Rosenbach 2005; Hinrichs and Szmrecsányi 2007; Grafmiller and Shih 2011; Shih et al., to appear):

a. [the attitude]$_{NP}$ of [people who are really into classical music and feel that if it's not seventy-five years old, it hasn't stood the test of time]$_{NP}$ >
b. [people who are really into classical music and feel that if it's not seventy-five years old, it hasn't stood the test of time]$_{NP}$'s [attitude]$_{NP}$

- Tagalog word order is usually VSO and, like English, heavy-last is expected (cf. Japanese).
  - e.g., Relative clause order is variable, but Schachter and Otanes (1972: 123) note a "tendency to prefer the order head-linker-modifier when the modifying phrase is long."

(25) ang pagkái-ng nilúto mo > ang nilúto mo-ng pagkáin (S&O 1972: 123)
DET food-LINK cooked you DET cooked you-LINK food 'the food you cooked'

➢ Tagalog word order: Results here, however, demonstrate a different effect: long words (with more segments) tend to prefer canonical, "default", adjective-noun order[2]. (Also tested syllable count with the same result.)

↳ Longer nouns prefer default second position.
(β = -0.001574; *SE* = 0.000445, t = -3.54)

(26) dakíla-ng kapangyaríhan > kapangyaríha-ng dakíla[?]
great-LINK power power-LINK great 'great power'

↳ Longer adjectives prefer default first position.
(β = -0.005807; *SE* = 0.00179, t = -3.23)

(27) pansamantalá-ng lúpon > lúpo-ng pansamantalá
temporary-LINK committee committee-LINK temporary 'temporary committee'

↳ Interaction for long nouns and long adjectives, even more "default" effect: longer nouns and adjectives prefer adjective-noun order.
(β = -0.000381, *SE* = 0.000209, t = -1.82)

(28) pangunáhi-ng katotohánan > katotohána-ng pangunáhin
basic-LINK fact fact-LINK basic 'basic facts'

- Possible explanation (for future inquiry):
  - In order to ease processing of heavy constituents, default order is used?

[2] cf. Donohue (2007:360, fn 11), who observes that "relative length" affects order.

5.3 NON-PHONOLOGICAL RESULTS

5.3.1 Prototypical adjectives

- Prototypical items are more easily accessed, and tend to come first in word order (Kelly et al. 1986).
- In Tagalog, adjectives are often formed from *ma-* prefix (e.g., *ma-baho* 'malodorous' from *baho* 'bad smell').

➢ Tagalog word order: *Ma-* initial adjectives—i.e., "prototypical" adjectives—prefer the predominant adjective-noun order.
(β = -0.012272, *SE* = 0.005007, t = -2.45)

(29)

	mabáho-ng	utót	>	utót	na	**ma**báho[?]
	malodorous-LINK	fart		fart	LINK	malodorous

5.3.2 Limiters

- Schacter and Otanes (1972:121): "Limiter adjectives" (e.g., cardinal numbers, ordinal numbers, quantifiers) prefer first position.
(β = -0.039683, *SE* = 0.018935, t = -2.1)

(30)

	lahát	na	táo	>	táo-ng	**lahát**	
	all	LINK	person		person-LINK	all	'all people'

5.3.3 Frequency

- More frequent nouns prefer initial position. (β = 0.00611, *SE* = 0.000772, t = 7.92).
- Frequency effect for adjectives not reliable (likely because they already prefer first position; also possible ceiling effect of frequency)[3].
- More frequent adjective/noun pairs prefer canonical, adjective-noun order[4].
(β = 0.051286, *SE* = 0.01049, t = 4.89)

[3] We also found a puzzling interaction between adjective frequency and pair: more frequent adjectives that appear in more frequent adjective/noun pairs (regardless of noun frequency) have a tendency to occur in noun-adjective order (β = 0.007679, *SE* = 0.003173, t = 2.42)

[4] Pair frequency (how often a given adjective and noun combination occurs) was residualized on both log(adjective frequency) and log(noun frequency).

6 DISCUSSION

- Our study results show that, controlling for non-phonological predictors, phonological conditions influence adjective/noun word order.
  - ↳ Word order alternations can act as repairs to satisfy phonological conditions (see also Schlüter 2005; Shih, in prep; a.o.).
- Similar effects of phonology influencing higher-level operations such as word order and construction choices have been demonstrated in other languages.
  - However, large-scale quantitative studies of these types of effects have mostly been limited to English.

6.1 PHONOLOGICAL EFFECTS ON WORD ORDER IDENTIFIED

- What type of phonological conditions can affect word order?
- Hypothesis (see also e.g., Shih, in prep): Phonological conditions that tend to affect word order are ones that hold over syntagmatic configurations—i.e., likely to trigger repairs when elements (e.g., words and phrases) (linearly) combine.
  - *Previously identified phonological effects on word order*
    Phonotactic, OCP
    Syllable structure optimization (e.g., hiatus avoidance)
    Prosody: rhythm, weight
  - *Newly identified phonological effects on word order*
    Contextual markedness (e.g., *NC̥)
    Long-distance phonotactics
    Phonological and morphological alignment

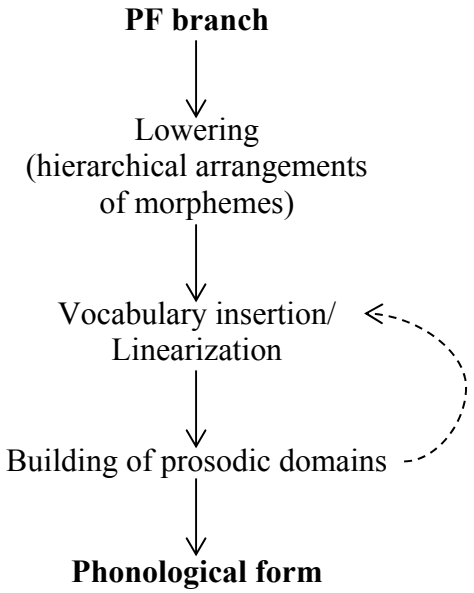
❖ (Syntagmatic) phonological and morphological conditions, which are more familiarly satisfied via phonological optimization, can also be satisfied via word order.

6.2 DISCUSSION: THEORETICAL IMPLICATIONS

- Syntax → Phonology generally have a feed-forward relationship in standard formal models of grammar (Zwicky & Pullum 1986) and psycholinguistic models of language production (Levelt 1989; Ferreira & Slevc 2007; cf. Vigliocco & Hartsuiker 2002)

- A more concrete example: in Distributed Morphology (DM), phonological exponence is not available to syntactic operations (late-insertion theory).
- Current question in DM: What is the ordering of linearization and vocabulary insertion, which provides phonological exponents? (see e.g., Embick 2007).
  - linearization prior to phonological information? (Embick & Noyer 2001; Pak 2008)
  - prosodic structure-building interleaved with linearization? (e.g., Tucker 2011)
  - more cycles through PF? (e.g., Embick 2007).

(31) see e.g., Embick and Noyer (2001: 566)



- Even if prosodic structure is allowed to interact with linearization, the general assumption is that segmental phonological information in the phonological form is largely unavailable and irrelevant for linearization purposes.

❖ **Open question:** How much phonological information is available at the point of linearization? How much phonological encoding occurs before/during grammatical encoding?

- The current study: looks at what happens when there are additional morphophonological processes that compound the problem of available phonological information during linearization.
- Phonologically-conditioned surface form of the linker particle must be available at the time of choice:
  - *-ng* particle preferred, and
  - segmental constraints (e.g., hiatus avoidance, OCP effects) make reference to the surface form.

❖ Suggests that effects of word order are in part dependent on knowledge and availability of surface segmental phonological information (not just prosodic structure).

- A possible alternative counter-hypothesis would be that the morphosyntax generates both options equally (i.e., both are viable grammatical alternatives) and then phonology filters.
- *However*, both orders do not appear to be equal options:
  - Adjective-first order is overwhelmingly preferred, even with controls for word, frequency, etc.
  - Phonological structure of adjective-noun order is more closely regulated: a greater number of robust constraints in the model penalize poor phonological structure in adjective-noun order than in noun-adjective order.

7 CONCLUSION

- Presented quantitative study of phonological and (some) non-phonological conditions on adjective/noun word order variation in Tagalog.

Results

- Though variable, adjective/noun pairs exhibit a canonical order preference for adjectives followed by nouns.
- Phonological conditions—both segmental and prosodic—influence adjective/noun word order preference.
- The phonologically-determined surface allomorph of the linker particle affects adjective/noun word order choice.

❖ Word order variation can be used to optimize phonological well-formedness, suggesting that phonological information factors into considerations of linearization, word order, and/or grammatical encoding.

Next directions

- Variable versus frozen adjective/noun pairs
  - Mollin (2012): Obeying well-formedness preferences that condition variable word order can cause word orders to become frozen → frozen word orders obey well-formedness preferences to a greater extent than their variable counterparts.
  - Do adjective/noun pairs in Tagalog that exhibit less reversibility demonstrate more adherence to the conditions on word order presented here?
  - May shed light on how phonological effects on word order may become grammaticalized over time.
- Variation with linker particle:
  - Variation in linker realization: bare adjective/noun pairs without linker are rare but do occur.
  - Variation in linker surface form: e.g., Richards (1999: 307) reports that some speakers accept linker *na* when *-ng* is called for phonologically, but this only occurs in noun-adjective order.
  - Open questions: What conditions such linker variation? and how does it interact with word order variation?

8 REFERENCES

Anttila, Arto, Matthew Adams & Michael Speriosu. 2010. The role of prosody in the English dative alternation. *Language and Cognitive Processes* 25(7-9). 946–981.

Bates, Douglas, Martin Maechler & Ben Bolker. 2013. *Package "lme4."* http://cran.r-project.org/web/packages/lme4/lme4.pdf.

Behagel, O. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25. 110–142.

Benor, Sarah Bunin & Roger Levy. 2006. The Chicken or the Egg? A Probabilistic Analysis of English Binomials. *Language* 82(2). 233–278.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & Harald Baayen. 2007. Predicting the dative alternation. In G Boume, I Kraemer & J Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Donohue, Mark. 2007. Word order in Austronesian from north to south and west to east. *Linguistic Typology* 11(2). 349–391.

Embick, David. 2007. Linearization and Local Dislocation: Derivational Mechanics and Interactions. *Linguistic Analysis* 33(3–4). 303–336.

Embick, David & Rolf Noyer. 2001. Movement Operations after Syntax. *Linguistic Inquiry* 32(4). 555–595.

English, Leo. 1986. *Tagalog-English dictionary*. Manila: Congregation of the Most Holy Redeemer; distributed by Philippine National Book Store.

Ferreira, Victor S. & L. Robert Slevc. 2007. Grammatical encoding. In Mark Gareth Gaskell (ed.), *The Oxford Handbook of Psycholinguistics*, 453–469. Oxford, UK: Oxford University Press.

French, Koleen Matsuda. 1988. *Insights into Tagalog: Reduplication, infixation, and stress from nonlinear phonology*. (Publications in Linguistics 84). Dallas, TX: The Summer Institute of Linguistics and The University of Texas at Arlington.

French, Koleen Matsuda. 1991. Secondary stress in Tagalog. *Oceanic Linguistics* 30(2). 157–178.

Frisch, Stefan A., Janet B. Pierrehumbert & Broe, Michael B. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22. 179–228.

Gelman, Andrew. 2008. Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine* 27. 2865–2873.

Gelman, Andrew & Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. (Analytic Methods for Social Research). New York, NY: Cambridge University Press.

Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In Y. Miyashita, A. Marantz & W. O'Neil (eds.), *Image, Language, Brain*, 95–126. Cambridge, MA: MIT Press.

Golston, Chris. 1995. Syntax Outranks Phonology: Evidence from Ancient Greek. *Phonology* 12(3). 343–368.

Grafmiller, Jason & Stephanie S Shih. 2011. New approaches to end weight. Paper. Paper presented at the Variation and Typology: New trends in Syntactic Research, Helsinki, Finland. http://stanford.edu/~stephsus/GrafmillerandShihHelsinki2011.pdf.

Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge, UK: Cambridge University Press.

Hayes, Bruce & Tanya Stivers. 1996. *The phonetics of post-nasal voicing*.

Hinrichs, Lars & Benedikt Szmrecsányi. 2007. Recent changes in the function and frequency of standard English genitive constructions: a multivariate analysis of tagged corpora. *English Language and Linguistics* 11(3). 437–474.

Kelly, Michael H, J. Kathryn Bock & Frank C Keil. 1986. Prototypicality in a linguistic context: Effects on sentence structure. *Journal of Memory and Language* 25(1). 59–74.

Kroeger, Paul. 1998. Nouns and verbs in Tagalog: a reply to Foley. *3rd LFG Conference*. Brisbane.

Levelt, Willem. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Löfstedt, Ingvar. 2010. Phonetic Effects in Swedish Phonology: Allomorphy and Paradigms. UCLA Ph.D. Dissertation.

McCarthy, John J. & Alan Prince. 1993. Generalized Alignment. *Yearbook of Morphology*. 79–154.

McDonald, Janet L., Kathryn Bock & Michael H. Kelly. 1993. Word and World Order: Semantic, Phonological, and Metrical Determinants of Serial Position. *Cognitive Psychology* 25. 188–230.

Menn, Lise & Brian MacWhinney. 1984. The Repeated Morph Constraint: Toward an Explanation. *Language* 60(3). 519–541.

Mollin, Sandra. 2012. Revisiting binomial order in English: ordering constraints and reversibility. *English Language and Linguistics* 16(1). 81–103.

Pak, Marjori. 2008. The postsyntactic derivation and its phonological reflexes. Philadelphia, Pennsylvania: University of Pennsylvania Ph.D. dissertation.

Pater, Joe. 1996. *NC. In Jill Beckman (ed.), *Proceedings of the North East Linguistics Society 26*, vol. 26, 227–239. Amherst, Mass.: GLSA Publications.

Pater, Joe. 2001. Austronesian nasal substitution revisited: What's wrong with *NC (and what's not). In Linda Lombardi (ed.), *Segmental Phonology in Optimality Theory: Constraints and Representations*, 159–182. Cambridge: Cambridge University Press.

Prince, Alan & Paul Smolensky. 1993. *Optimality Theory*. Blackwell.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London and New York: Longman.

Richards, Norvin. 1999. Complementizer cliticization in Tagalog and English. *Proceedings of AFLA VI*, vol. 16(2), 297–311. (Toronto Working Papers in Linguistics). https://jps.library.utoronto.ca/index.php/twpl/article/view/6294.

Rosenbach, Anette. 2005. Animacy versus Weight as Determinants of Grammatical Variation in English. *Language* 81(3). 613–644.

Schachter, Paul & Fe T Otanes. 1972. *Tagalog Reference Grammar*. Berkeley, CA: University of California Press.

Schlüter, Julia. 2005. *Rhythmic Grammar: The Influence of Rhythm on Grammatical Variation and Change in English*. Berlin, Germany: Mouton de Gruyter.

SEAsite. 2001. SEAsite Tagalog Online Dictionary (Large Tagalog Dictionary version). http://www.seasite.niu.edu/Tagalog/on_site_dictionary.htm.

Shih, Stephanie S. in prep. Towards Optimal Rhythm. Stanford, CA: Stanford University PhD dissertation.

Shih, Stephanie S, Jason Grafmiller, Richard Futrell & Joan Bresnan. to appear. Rhythm's role in predicting genitive alternation choice in spoken English. In Ralf Vogel & Ruben van de Vijver (eds.), *Rhythm in phonetics, grammar, and cognition*.

Stemberger, Joseph. 1981. Morphological haplology. *Language* 57. 791–817.

Szmrecsányi, Benedikt. 2004. On Operationalizing Syntactic Complexity. *7es Journ´ees internationales d'Analyse statistique des Donn´ees Textuelles*, 1031–1038. http://www.benszm.net/omnibuslit/Szmrecsanyi2004.pdf.

Temperley, David. 2006. Minimization of dependency length in written English. *Cognition* 105. 300–333.

Temperley, David. 2009. Distributional Stress Regularity: A Corpus Study. *Journal of Psycholinguistic Research* 38(1). 75–92.

Tucker, Matthew A. 2011. The Morphosyntax of the Arabic Verb: Toward a Unified Syntax-Prosody. *Morphology at Santa Cruz: Papers in Honor of Jorge Hankamer*, 177–211. Santa Cruz, CA: Linguistics Research Center, University of California, Santa Cruz.

Vigliocco, Gabriella & Robert J. Hartsuiker. 2002. The Interplay of Meaning, Sound, and Syntax in Sentence Production. *Psychological Bulletin* 128(3). 442–472.

Wasow, Tom. 2002. *Postverbal Behavior*. Stanford, CA: CSLI Publications.

Wright, Saundra K., Jennifer Hay & Tessa Bent. 2005. Ladies first? Phonology, frequency, and the naming conspiracy. *Linguistics* 43(3). 531–561.

Yip, Moira. 1998. Identity avoidance in phonology and morphology. In Steven G Lapointe, Diane K Brentari & Patrick M Farrell (eds.), *Morphology and its Relation to Phonology and Syntax*, 216–246. Stanford, CA: CSLI Publications.

Zec, Draga & Sharon Inkelas. 1990. Prosodically constrained syntax. In Sharon Inkelas & Draga Zec (eds.), *The Phonology-Syntax Connection*, 365–378. Chicago: University of Chicago Press.

Zubizarreta, Maria Luisa. 1998. *Prosody, Focus, and Word Order*. (Linguistic Inquiry Monographs 33). Cambridge, MA: MIT Press.

Zuraw, Kie. 2006. Using the Web As a Phonological Corpus: A Case Study from Tagalog. *Proceedings of the 2Nd International Workshop on Web As Corpus*, 59–66. (WAC '06). Stroudsburg, PA, USA: Association for Computational Linguistics. http://dl.acm.org/citation.cfm?id=1628297.1628306

Zuraw, Kie. 2010. A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language & Linguistic Theory* 28(2). 417–472.

Zwicky, Arnold M. 1987. Suppressing the Z's. *Journal of Linguistics* 23. 133–148.

Zwicky, Arnold M. & Geoffrey Pullum. 1986. The principle of phonology-free syntax: introductory remarks. *Interfaces*, vol. 32, 63–91. (Ohio State University Working Papers in Linguistics).